I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0205

The role of the exogenous randomness in the Identification of Conditional Models

A. Oulhaj and M. Mouchart

http://www.stat.ucl.ac.be

The role of the exogenous randomness in the Identification of Conditional Models

A. Oulhaj[1,2] and M. Mouchart[2]

Institut de statistique, Université catholique de Louvain.

Abstract

This paper draws the attention on several subtleties requiring a proper approach to both conditional models and their identification. We start with a definition of identification in conditional models called *weak* identification, derived from the usual concept of identification in standard statistical models. We show, in theorem 1, that weak identification coincides with that one usually met in the statistical literature. Theorem 4 shows, however, an undesirable property of weak identification, namely that under rather general conditions, the weak identification does not depend on the sample size. An example of a standard regression model is given to illustrate why such property is undesirable. As an alternative, three other levels of identification are given, stressing the role of the randomness of the conditioning variables. Similar distinctions are also shown to be relevant for properties of estimators, such as unbiasedness or consistency. The relation between these different levels of identification, unbiasedness and consistency are given in theorems 2, 3, 5 and 6.

Keywods: Identification; Consistency; Unbiasedness; Conditional models.

---

[1]Corresponding author. A. Oulhaj, Institut de statistique, 20 voie du roman pays, 1348 LLN, Belgium; e-mail: oulhaj@stat.ucl.ac.be

[2]Support of the contract *Projet d'Actions de Recherche Concertées* ARC 98/03-217, of the Belgium Governement, is gratefully acknowledged.

1 Introduction

The object of this paper is to make explicit the role of the exogenous randomness in the identification of conditional models. The starting idea is to give a general definition of a conditional model through embedding that concept into the concept of a standard statistical model, and to analyze the identification of a conditional model in the framework of the identification of a function of the parameters in a standard statistical model. Taking explicitly into account the randomness of the exogenous variables, say Z, we show that the identification of conditional models may crucially depend on conditions on Φ, the set of distributions of Z.

This paper gives solutions to three issues in statistical modelling. First, contrary to standard models, identification in conditional models may not be considered as a unique concept. Several levels should be distinguished, according to the requirements made in terms of the distribution of the exogenous variables. Secondly, similar distinctions are shown to be also required when defining properties of estimators in conditional models, such as unbiasedness or consistency. Thirdly, this paper makes the relationships between these different levels of concept explicit. The main practical implication is that, either identification in conditional model is controlled at a stronger level in order to ensure lower levels, or identification is established at a particular level in order to ensure the properties of an estimator at a particular level of specific interest.

The paper is organized as follows: In section 2, we briefly review the identification condition in the standard statistical models. Section 3 treats the identifcation in conditional models; we first briefly sketch the structure of binary response models drawn from Manski (1988) and give his identification condition. Next, we start with a definition of identification in conditional models called *weak* identification, derived from the concept of identification in standard statistical models. We show, in theorem 1, that weak identification coincides with that one usually met in the statistical literature; in particular in Manski (1988). Furthermore, theorem 4 shows an undesirable property of weak identification, namely that under rather general conditions, the weak identification does not depend on the sample size. An example of a standard regression model is given to illustrate why such a property is undesirable. As an alternative, three levels of identification are given in section 4, stressing the proper role of the randomness of the conditioning variables. Similar distinctions are also shown to be relevant for properties of estimators, such as unbiasedness or consistency. The relationships between these different levels of identification, unbiasedness and consistency are given . The final section gives concluding remarks on the practical relevance of these distinctions and also gives a graphical scheme illustrating the links between the different concepts. An appendix collects the proofs of the main results.

2 Basic Concepts: Identification in standard statistical models

Let us start with a *standard parametrised* statistical Model $\mathbf{M}^{\omega}_{\mathbf{X}}$ given in the following form:

$$\mathbf{M}^{\omega}_{\mathbf{X}} \;=\; [(\mathbf{R}_X\ ,\ \mathcal{X})\ ,\ \{P^{\omega}\ :\ \omega\ \in\ \Omega\}] \tag{1}$$

where $(\mathbf{R}_X\ ,\ \mathcal{X})$ is the sample space of X, P^{ω} is a probability measure on $(\mathbf{R}_X\ ,\ \mathcal{X})$ for each $\omega \in \Omega$ and Ω is the parameter space. For the sake of simplicity, we adopt the following notation across this paper: For any function $h : E \longrightarrow F$ we define $E_h = \{(x_1, x_2) \in E^2 :\ h(x_1) \neq h(x_2)\}$, *i.e.* E_h is the set of pairs (x_1, x_2) separated by h. We also consider $g(\omega)$ as a function defined on the parameter space and representing the parameter of interest.

Let us recall the usual concept of identification in this framework.

Definition 1 *A parameter function* $g(\omega)$ *is said to be identified if:*

$$\forall\,(\omega_1, \omega_2) \in \Omega_g :\ \ P^{\omega_1} \neq P^{\omega_2} \tag{2}$$

The *model* is *identified* if the identity function $g(\omega) = \omega$ is identified, *i.e.* if:

$$\forall\,\omega_1 \neq \omega_2 :\ \ P^{\omega_1} \neq P^{\omega_2} \tag{3}$$

Thus a better expression would be that the *parametrization* ω , rather than the model, is *identified.* As a consequence from now on, we only consider the identification of functions of ω explicitly, leaving implicit the definition of the identification of a model as corresponding to the identification of the identity function. Clearly, ω is identified implies that $g(\omega)$ is identified for any function g. For more details on identification see, among others, Bunke and Bunke (1974), Florens et al (1985), Koopmans and Reiersøl (1950) and Roehrig (1988).

3 Identification in conditional models

3.1 A look on the literature

In the statistical literature, these authors have found no general definition designed specifically for the identification in conditional models. However, we have found for some specific models, conditions of identification of their parameters. As an example, we briefly sketch the structure of binary reponse models drawn from Manski (1988) and give his identification condition.

Let $X = (Y, Z)$ be an observable vector where $Y \in \{0, 1\}$ is an endogenous variable representing the individual observed choice, and $Z \in \mathbf{R}_Z \subset I\!R^k$ is a vector of exogenous variables. Let also $\zeta \in I\!R$ be a latent variable connected with the observed choice Y by the relation: $Y = \mathbb{1}_{\{Z'\beta+\zeta>0\}}$, where $\mathbb{1}_{\{.\}}$ is the indicator function. In a semi-parametric framework, the parameter characterizing the structural model is $\theta = (\beta, F_\zeta^Z) \in \Theta \subset I\!R^k \times \mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ where $\mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ is the set of all transitions defined on $\mathbf{R}_Z \times \mathcal{B}$, with $\mathcal{B}$ being the Borel σ-field of $I\!R$. The statistical model describing the Z-conditional observed choice probabilities is deduced from the structural model as follows:

$$I\!E(Y \mid Z, \theta) = P(Y = 1 \mid Z, \beta, F_\zeta^Z) = 1 - F_\zeta^Z(-Z'\beta)$$

Following Manski (1988), the prameter $\theta = (\beta, F_\zeta^Z)$ is *identified* relative to another parameter $\theta' = (b, G_\zeta^Z)$ if the Z-conditional observed choice probabilities generated by θ and θ' are different over an F_Z-non nul subset of $\mathbf{R}_Z$, where F_Z represent a *non specified* distribution of the exogenous vector Z; more specifically:

Definition 2 *(Manski 1988) Let $B(\theta, \theta') = \{z \in \mathbf{R}_Z : I\!E(Y \mid z, \theta) \neq I\!E(Y \mid z, \theta')\}$. The parametrization θ is said to be identified if $\forall\, \theta' \neq \theta$: $F_Z(B(\theta, \theta')) > 0$. Furthermore, a parameter of interest $g(\theta)$ is identified if $\forall\, g(\theta') \neq g(\theta) :\ F_Z(B(\theta, \theta')) > 0$.*

Clearly, this definition of identification crucially depends on the distribution F_Z of the exogenous variables Z. As a matter of fact, Manski (1988) gives conditions under wich the complete parameter, say $\theta = (\beta, F_{\zeta}^{Z})$ (or a parameter of interest such as $g(\theta) = \beta$, $g(\theta) = sign(\beta)$, ...) is identified. These conditions depend, among other, on specific assumptions on F_Z.

3.2 Conditional models

Let us now construct the conditional model through embedding that concept into the usual concept of a standard statistical model (1). For the sake of simplicity, this section only considers the case where a random vector X of observations is decomposed into $X' = (Y', Z')$ (where ' denotes transposition) with values in $(\mathbf{R}_Y \times \mathbf{R}_Z\ ,\ \mathcal{Y} \otimes \mathcal{Z})$ and the model is conditional on Z; more general conditioning might be introduced at the cost of a somewhat more abstract presentation but without substantial benefits for the object of this paper.

The basic idea of a conditional model $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ is the following : starting from a global model $\mathbf{M}_{\mathbf{X}}^{\omega}$, each sampling probability P is first decomposed into the Markovian product:

$$P \ = \ P_Z \otimes P_Y^Z$$

where P_Z is the marginal probability on $(R_Z\ ,\ \mathcal{Z})$ and P_Y^Z is the conditional probability given $\mathcal{Z}$; more explicitly, P_Y^Z is defined by means of the identity:

$$\forall (A, B) \in \mathcal{Y} \times \mathcal{Z}\ :\ P(A \times B) \ = \ \int_B P_Y^Z(A)\, dP_Z \tag{4}$$

Next, one makes specific assumptions on the conditional component leaving virtually unspecified the marginal component. Thus a conditional model $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ may be represented as follows :

$$\begin{aligned}\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta} \ = \ [(\mathbf{R}_Y \times \mathbf{R}_Z\, , \mathcal{Y} \otimes \mathcal{Z}) \quad , \quad & \{P^{\omega} \ = \ \phi \otimes P_Y^{Z,\theta} : \\ & \omega = (\theta, \phi) \in \Omega = (\Theta \times \Phi)\}]\end{aligned} \tag{5}$$

where Φ is a set of, possibly all, probability measures on $(\mathbf{R}_Z, \mathcal{Z})$ and for each $\theta \in \Theta, P_Y^{Z,\theta}$ is a transition of probability defined on $\mathbf{R}_Z \times \mathcal{Y}$, with values denoted as $P_Y^{z,\theta}(A)$ for any $(z, A) \in (\mathbf{R}_Z \times \mathcal{Y})$. The essential features of a conditional model are therefore:

(i) θ indexes a "well specified" family of transitions of probability defined on $\mathbf{R}_Z \times \mathcal{Y}$. This family constitutes the kernel of the concept of a conditional model. This concept relates, however, to a family of joint probabilities (on $\mathcal{Y} \otimes \mathcal{Z}$) by crossing the family of transitions with a family of probabilities on $\mathcal{Z}$. By doing so, for each $\omega = (\theta, \phi)$, the transition $P_Y^{Z,\theta}$ becomes a particular version of the conditional probability given $\mathcal{Z}$, for each ϕ defining a joint probability $\phi \otimes P_Y^{Z,\theta}$.

(ii) θ is the only parameter of interest and ϕ is a nuisance parameter which is identified by definition (because $\phi \in \Phi \subset \mathcal{P}(\mathbf{R}_Z, \mathcal{Z})$). Furthermore θ and ϕ are variation free. The notation $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ conveys the idea that θ is the only actual parameter of interest, leaving to ϕ no explicit role.

(iii) The modelling restrictions are concentrated on the conditional part *i.e.* the set $\{P^{Z,\theta} : \theta \in \Theta\}$ embodies the main hypotheses of the model, whereas in most cases, the set Φ embodies a minimal amount of restrictions, typically only the hypotheses necessary to guarantee essential properties for the inference on θ, such as identifiability or convergence of estimators. Consequently, in most situations, but not in all, Φ is a "thick" subset of the set of all probability measures on $(R_Z, \mathcal{Z})$. We shall see nevertheless that Φ may play an important role. This will be illustrated in next sections.

(iv) In an asymptotic model, $X = (Y, Z)$ correspond to infinite sequences, *i.e.* to a stochastic process in discrete time, $Y = (Y_1, Y_2, ..., Y_i, ...)$ and $Z = (Z_1, Z_2, ..., Z_i, ...)$. Finite sequences are denoted as $X_1^n = (Y_1^n, Z_1^n)$ with $Y_1^n = (Y_1, Y_2, ..., Y_n)$ and similarily for Z_1^n. The sample space $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$ is decomposed into pieces corresponding to samples of size 1, namely $(\mathbf{R}_{Y_i} \times \mathbf{R}_{Z_i}, \mathcal{Y}_i \otimes \mathcal{Z}_i)$. For every $\omega = (\theta, \phi) \in \Omega = \Theta \times \Phi$, the law $P^\omega = \phi \otimes P_Y^{Z,\theta}$ of the stochastic process $X = (Y, Z)$ induces a sequence of marginal probability measures $P^{\omega^{(n)}} = \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta^{(n)}}$ on $X_1^n = (Y_1^n, Z_1^n)$. The parameter space of the conditional model (5) is accordingly decomposed into $\omega^{(n)} = (\phi^{(n)}, \theta^{(n)}) \in \Phi^{(n)} \times \Theta^{(n)}$. In many instances, but not always, $\Theta^{(n)} = \Theta$ for any $n \geq 1$.

3.3 Weak identification

In this subsection, we introduce a concept of *weak* identification, as a natural extension of the concept usual in a standard statistical model. We also show that this concept has been used in the statistical literature but is liable to substantial drawbacks, to be corrected in the next section.

The parameter function $g : \Theta \longrightarrow \Gamma$ is identified from the *joint* model in (5) if and only if the parameter function $g \circ h$ is identified from that model, where $h : \Theta \times \Phi \longrightarrow \Theta$ such that $h(\theta, \phi) = \theta$. The identification of $g \circ h$ in (2):

$$\forall\, (\omega_1, \omega_2) \in \Omega_{g \circ h} : \; P^{\omega_1} \neq P^{\omega_1} \tag{6}$$

is equivalent to:

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall (\phi_1, \phi_2) \in \Phi^2 : \; \phi_1 \otimes P_Y^{Z,\theta_1} \; \neq \; \phi_2 \otimes P_Y^{Z,\theta_2} \tag{7}$$

As the nuisance parameter ϕ is identified, equation (6) or (7) may, equivalently, be written in the form of the following definition.

Definition 3 *The function $g(\theta)$ is weakly identified if:*

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi \in \Phi : \; \phi \otimes P_Y^{Z,\theta_1} \; \neq \; \phi \otimes P_Y^{Z,\theta_2} \tag{8}$$

Weak identification is accordingly the direct transposition of the usual identification in a complete model to the case of conditional model, and is explicitely based on the joint distribution on $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$, *i.e.* $\phi \otimes P_Y^{Z,\theta}$. Weak identification satisfies usual properties of identification in standard statistical model, namely:

1) If g is weakly identified, then $k \circ g$ is also weakly identified for any function k defined on $g(\Theta)$.

2) If g is weakly identified in $\mathbf{M}_{Y_1^{n_0}}^{Z_1^{n_0},\theta}$, then g is weakly identified in $\mathbf{M}_{Y_1^n}^{Z_1^n,\theta}$ for any $n \geq n_0$.

3) Let $U_n = f_1(X_1^n)$ and $V_n = f_2(X_1^n)$ be two statistics. If g is identified in $\mathbf{M}_{V_n}^{Z_1^n,\theta}$, then g is also identified in $\mathbf{M}_{U_n}^{Z_1^n,\theta}$ for any $f_2 = k \circ f_1$ with k defined on $f_1(\mathbf{R}_{Z_1^n})$.

Next theorem gives a suitable assumption, under which an equivalent and more operational condition for the weak identification is obtained.

Theorem 1 *For any* $(\theta_1, \theta_2) \in \Theta^2$ *let* $B(\theta_1, \theta_2) = \{z \in \mathbf{R}_Z \mid \ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}$. *If* $\mathcal{Y}$ *is separable (i.e. countably generated), then:*

(i) $B(\theta_1, \theta_2)$ *is* $\mathcal{Z}$*-measurable for any* $(\theta_1, \theta_2) \in \Theta^2$.

(ii) A necessary and sufficient condition for the weak identification of $g(\theta)$ *is:*

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi \in \Phi: \ \phi(B(\theta_1, \theta_2)) > 0. \tag{9}$$

The proof of theorem 1 is essentially based on the following lemma.

Lemma 1 *If* $\mathcal{Y}$ *is separable, the following two statements are equivalent:*

$$\phi \otimes P_Y^{Z,\theta_1} = \phi \otimes P_Y^{Z,\theta_2} \tag{10}$$

$$P_Y^{Z,\theta_1} = P_Y^{Z,\theta_2} \quad \phi - a.s \tag{11}$$

Theorem 1 shows that, under a separability condition, weak identification coincides with a concept of identification used in the statistical literature; in particular, for binary response Model as seen in Manski (1988) and for a polychotomous choice models as seen in Matzkin (1993). The separability of $\mathcal{Y}$ ensures the measurability of the set $B(\theta_1, \theta_2)$ used in theorem 1, and eventually solves an issue raised in Manski (1988). Note that when $(\mathbf{R}_Y, \mathcal{Y})$ is euclidean (and $\mathcal{Y}$ is the usual Borel σ-field), separability is ensured; however when Y is treated as a functional observation, as in electrocardiograms or in human or animal growth curves, care should be taken to ensure the separability of $\mathcal{Y}$, typically by endowing $\mathbf{R}_Y$ with a separable topology and taking for $\mathcal{Y}$ the corresponding Borel σ-field.

3.3.1 Weak identification, weak unbiasedness and weak consistency

Next Theorem shows the relation between *weak* identification and *weak* unbiasedness. Let us first introduce the concept of *weak* unbiasedness.

Definition 4 *Let T be an estimator of $g(\theta)$, i.e. $T = f(X)$. Then, T is said to be weakly unbiased for $g(\theta)$ if:*

$$\forall\,(\theta, \phi) \in \Theta \times \Phi : \;\; I\!E^{\theta,\phi}(T) = g(\theta) \tag{12}$$

Weak unbiasedness is accordingly the direct transposition of the usual unbiasedness in a complete model to the case of conditional model and is explicitely based on the joint distribution on $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$, *i.e.* $\phi \otimes P_Y^{Z,\theta}$. Similarly to the weak identification, the *unique* conditional aspect of the concept of weak unbiasedness is that T estimates a function of the parameter θ characterising the conditional distribution.

Theorem 2 *A necessary condition for the existence of a weakly unbiased estimator of $g(\theta)$ is that $g(\theta)$ is weakly identifiable.*

The proof of theorem 2 is an immediate consequence of the fact that, if T is *weakly* unbiased for $g(\theta)$ then:

$$\forall\,(\theta_1, \theta_2) \in \Theta_g, \forall\,\phi \in \Phi : \;\; I\!E^{\theta_1,\phi}(T) \neq I\!E^{\theta_2,\phi}(T) \tag{13}$$

which also implies that:

$$\forall\,(\theta_1, \theta_2) \in \Theta_g, \forall\,\phi \in \Phi : \;\; \phi \otimes P_Y^{Z,\theta_1} \neq \phi \otimes P_Y^{Z,\theta_2}$$

■

Similarly to definition 4, we now consider the concept of *weak* consistency in conditional models.

Definition 5 *Let T_n be a sequence of estimators of $g(\theta)$ (i.e. $T_n = f_n(X_1^n)$). Then, T_n is said to be weakly consistent if:*

$$\forall\,\omega = (\theta, \phi) \in \Theta \times \Phi, \forall\,\epsilon > 0 : \lim_{n \longrightarrow +\infty} P^{\omega}(|T_n - g(\theta)| < \epsilon) = 1 \tag{14}$$

This will be denoted $p^{\omega}lim\, T_n = g(\theta)$.

In this definition, P^w represents the law of the stochastic process $X = (Y, Z)$ (see property (iv) of conditional model, section 2.1). Again, the unique conditional aspect of the weak consistency is reduced to the fact that the parameter of interest, $g(\theta)$, is a function of θ which characterises the conditional distribution. Next theorem gives the link between weak identification and weak consistency.

Theorem 3 *A necessary condition for the existence of a weakly consistent estimator for $g(\theta)$ is that $g(\theta)$ is weakly identified in the asymptotic model.*

3.3.2 Weak identification and the sample size

In the *i.i.d* case of the standard statistical model (1), it is known that the identification does not depend on n, the sample size, in the sense that identification for $n = 1$, for finite n and in the asymptotic model are equivalent. It is accordingly of interest to consider the relationship between identification in finite sample and identification in the asymptotic model for conditional models. Next theorem states that under some rather general conditions, weak identification does not depend on the sample size. The first two conditions of this theorem may be viewed as "convenience assumptions", they are not substantially restrictive but facilitate the exposition:

(C.1) $\forall\, i \geq 1: \quad (\mathbf{R}_{Y_i} \times \mathbf{R}_{Z_i}\, , \mathcal{Y}_i \otimes \mathcal{Z}_i) = (\mathbf{R}_{Y_1} \times \mathbf{R}_{Z_1}\, , \mathcal{Y}_1 \otimes \mathcal{Z}_1)$

(C.2) $\forall\, n \geq 1: \ \Theta^{(n)} = \Theta$

The next one is a coherence condition, similar, for a familiy of stochastic processes, to the Kolmogorov coherence condition (of a projective system) for a unique stochastic process. More specifically, let us define Φ^n_{n+1}, a set of transitions ϕ^n_{n+1} defined on $(\mathbf{R}_{Z_1^n} \times \mathcal{Z}_{n+1})$ (representing the distributions of Z_{n+1} given Z_1^n); this condition requires the sequence $\Phi^{(n)}$ to be monotonically increasing in the following sense:

(C.3) $\forall\, n \geq 1: \ \Phi^{(n+1)} = \{\phi^{(n)} \otimes \phi^n_{n+1} \mid \phi^{(n)} \in \Phi^{(n)}\, ; \phi^n_{n+1} \in \Phi^n_{n+1}\}$

Next condition requires that the set Φ^n_{n+1}, defined in (C.3), contains all the initial distributions $\phi^{(1)}$ understood as degenerate transitions, and eventually assumes that the family $\Phi^{(n)}$ contain the *i.i.d* case for the Z_i's.

(C.4) $\forall\, n \geq 1, \ \forall\, \phi^{(1)} \in \Phi^{(1)}, \ \exists\, \phi^n_{n+1} \in \Phi^n_{n+1}$ such that:

$$\forall\, \phi^{(n)} \in \Phi^{(n)}: \ \phi^{(n)}(\{z_1^n \mid \forall\, B \in \mathcal{Z}_1: \ \phi^n_{n+1}(B \mid z_1^n) = \phi^{(1)}(B)\}) = 1$$

The following and main condition is a sifting condition, which is standard in asymptotic theory (for details, see *e.g.* Florens et al. (1990), section 7.6), *viz.*

(C.5) $\perp\!\!\!\perp Y_i \mid Z_1^n, \theta$ and $Y_i \perp\!\!\!\perp Z_1^n \mid Z_i, \theta; \ \ \forall\, 1 \leq i \leq n.$

The last condition is, under (C.1) and (C.5), a condition of conditional stationarity:

(C.6) $\forall\, n \geq 1, \forall\, \theta \in \Theta: \; P_{Y_n}^{Z_n=z,\theta} = P_{Y_1}^{Z_1=z,\theta} \quad \forall\, z \in \mathbf{R}_{Z_1}$

Theorem 4 *Consider a parameter function $g : \Theta \longrightarrow \Gamma$. Under the assumptions (C.1) to (C.6), the following two relations are equivalent:*

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi^{(n+1)} \in \Phi^{(n+1)}: \; \phi^{(n+1)} \otimes P_{Y_1^{n+1}}^{Z_1^{n+1},\theta_1} \neq \phi^{(n+1)} \otimes P_{Y_1^{n+1}}^{Z_1^{n+1},\theta_2} \tag{15}$$

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi^{(n)} \in \Phi^{(n)}: \; \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta_1} \neq \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta_2} \tag{16}$$

Theorem 4 shows a characteristic property of weak identification, namely that under rather general conditions, the weak identification of a model does not depend on the sample size. Next example shows why such a property is eventually undesirable.

Example 1

Let us consider a standard linear regression model defined as follows $(Y|\mathbf{Z}, \beta, \sigma^2) \sim N_{(n)}(\mathbf{Z}\beta, \sigma^2 I_n)$, where $\mathbf{Z} \in I\!R^{n\times(k+1)}$ is a matrix of k exogenous variables plus a constant term and $\beta = (\alpha, \gamma')' \in I\!R^{k+1}$ ($\alpha \in I\!R$) is a vector of regression coefficients. Thus, $\theta = (\beta, \sigma^2) \in I\!R^{k+1} \times I\!R_+$. One can easily show that $g(\theta) = \sigma^2$ is weakly identified for any sample size. Let us now show under what conditions $g(\theta) = \beta$ is weakly identified. Taking into account the constant term in the regression, we first decompose the matrix $\mathbf{Z}$ as follows $\mathbf{Z} = (i_n, \mathbf{Z}_2)$, with $i_n = (1, 1, ..., 1)' \in I\!R^n$ and $\mathbf{Z}_2$ is a random $n \times k$ matrix of exogenous (random) variables. Following theorem 1, $g(\theta) = \beta$ is *weakly* identified if:

$$\forall\, \beta_1 \neq \beta_2, \forall\, \phi \in \Phi: \; \phi\{\mathbf{z} \in I\!R^{n\times(k+1)} \mid N(\mathbf{z}\beta_1\,,\, \sigma^2 I_n) \neq N(\mathbf{z}\beta_2\,,\, \sigma^2 I_n)\} > 0 \tag{17}$$

Let $\beta_1 = (\alpha_1, \gamma_1)$ and $\beta_2 = (\alpha_2, \gamma_2)$ and define:

$$B(\beta_1, \beta_2) = \{\mathbf{z}_2 \in I\!R^{n\times k} \mid \mathbf{z}_2(\gamma_2 - \gamma_1) \neq i_n(\alpha_1 - \alpha_2)\}$$

Equation (17) may equivalently be written as:

$$\forall\, \beta_1 \neq \beta_2, \forall\, \phi \in \Phi: \; \phi(B(\beta_1, \beta_2)^c) < 1 \tag{18}$$

A standard distributional assumption for the exogenous variables $\mathbf{Z}_2$ is to take $\Phi = \{\phi \mid \ \phi \ll \lambda_{n\times k}\}$ with $\lambda_{n\times k}$ being the Lebesgue measure on $I\!R^{n\times k}$. In such a case, the set $B(\beta_1, \beta_2)^c$ has zero lebesgue measure, because the dimension of $B(\beta_1, \beta_2)$ is equal to $n(k-1)$, and therefore has zero probability measure for every $\phi \in \Phi$. This make the regression coefficients of the standard regression model weakly identified even for a sample of size 1.

■

This example illustrates an unpalatable property of the weak identification because, for instance, in the case of a linear multiple regression model, the existence of a (linear) unbiased estimator of β requires the rank of the matrix $\mathbf{Z}$ to be equal to $k+1$ and therefore $n \geq k+1$, a seemingly natural requirement of identification. Theorem 4 suggests therefore that a concept of identification stronger than the weak one is called for. This is the object of next section.

4 Alternative definitions

4.1 Three levels of identification

The weak identification has an undesirable property, namely that the probability of values z of $\mathbf{R}_Z$ such that $g(\theta)$ would be identified in the model $\mathbf{M}^{\theta}_{Y|z} = \{(\mathbf{R}_Y, \mathcal{Y}), \{P^{z,\theta}_Y : \ \theta \in \Theta\}\}$ can be zero for every $\phi \in \Phi$. This leads us to introduce alternative definitions of identification in conditional models, taking better into account the randomness of the exogenous variables.

Next definition provides a natural strengthening of weak identification, at the cost of introducing some regularity conditions (paid for the sake of avoiding irrelevant complications). For this purpose, let us denote the set of z-values providing the identification of $g(\theta)$ in $\mathbf{M}^{\theta}_{Y|z}$ as follows:

$$\begin{aligned} \mathbf{R}_{Z,g} &= \{z \in \mathbf{R}_Z \mid \ \forall\, (\theta_1, \theta_2) \in \Theta_g : \ P^{z,\theta_1}_Y \neq P^{z,\theta_2}_Y\} \qquad (19) \\ &= \bigcap_{\{(\theta_1,\theta_2)\in\Theta_g\}} B(\theta_1, \theta_2) \end{aligned}$$

Definition 6 *Let* $\Phi \subset \mathcal{P}(\mathbf{R}_Z, \mathcal{Z})$ *and assume that* $\mathbf{R}_{Z,g} \in \mathcal{Z}$.
a) The parameter function $g(\theta)$ *is said to be identified in the conditional model* $\mathbf{M}_Y^{Z,\theta}$ *if* $\forall\, \phi \in \Phi : \ \phi(\mathbf{R}_{Z,g}) > 0$.

b) The parameter function $g(\theta)$ *is said to be almost surely identified in the conditional model* $\mathbf{M}_Y^{Z,\theta}$ *if* $\forall\, \phi \in \Phi : \ \phi(\mathbf{R}_{Z,g}) = 1$.

c) The parameter function $g(\theta)$ *is said to be* $\mathbf{R}_Z^*$*-uniformly identified in the conditional model* $\mathbf{M}_Y^{Z,\theta}$ *if* $\forall\, z \in \mathbf{R}_Z^*, \forall\, (\theta_1, \theta_2) \in \Theta_g : \ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}$. *When* $\mathbf{R}_Z^* = \mathbf{R}_Z$, *we say that* $g(\theta)$ *is uniformly identified.*

$\blacksquare$

Clearly, $g(\theta)$ is *uniformly* identified implies that $g(\theta)$ is *almost surely* identified which also implies that $g(\theta)$ is *identified* which also implies that $g(\theta)$ is *weakly* identified. The case where the exogenous variables are treated as if they were not random, corresponds to the case where Φ contains point masses only, *i.e.* $\Phi \subset \Phi_\delta$ where $\Phi_\delta = \{\delta_z : z \in \mathbf{R}_Z\}$ with δ_z denoting the point mass distribution (giving probability 1 to the point z), *i.e.* $\forall\, B \in \mathcal{Z} : \delta_z(B) = 1 \iff z \in B$. Uniform identification may be viewed as weak identification for the case $\Phi = \Phi_\delta$. Clearly, in such a case, the four concepts of identification coincide.

Example 2 (example 1 continued)

As an example, the regression model introduced above is identified (resp. almost surely identified) if:

$$\forall\, \phi \in \Phi : \ \phi\{\mathbf{z} \mid \ \forall\, \beta_1 \neq \beta_2 : \mathbf{z}(\beta_1 - \beta_2) \neq 0\} > 0 \ \ (\text{resp.} = 1) \tag{20}$$

Using a similar argument as above, we define:

$$\mathbf{R}_{Z_2,g} = \{\mathbf{z}_2 \in I\!R^{n\times k} \mid \forall\, \beta_1 \neq \beta_2 : \mathbf{z}_2(\gamma_2 - \gamma_1) \neq i_n(\alpha_1 - \alpha_2)\} \tag{21}$$

Equation (20) may equivalently be written as:

$$\forall\, \phi \in \Phi : \ \phi(\mathbf{R}_{Z_2,g}) > 0 \ \ (\text{resp.} = 1) \tag{22}$$

Which is also equivalent to:

$$\forall\, \phi \in \Phi : \ \phi\{\mathbf{z}_2 \in I\!R^{n\times k} \mid \ r(\mathbf{z}_2) = k\} > 0 \ \ (\text{resp.} = 1) \tag{23}$$

Where $r(\mathbf{z}_2)$ is the rank of the matrix $\mathbf{z}_2$.

In what follows, we give an example showing that the level of identification of β crucially depends on the specification of Φ. For this purpose, let us first partition the matrix $\mathbf{Z}_2$ according to its rows $Z_i, i = 1, 2, .., n$. Let us also assume that the Z_i's are mutually independent and identically distributed with common distribution ϕ, expectation μ_ϕ and covariance matrix Σ_ϕ and define:

$$\Phi_1^{iid} = \{\phi \mid \ \phi \ll \lambda_k\} \tag{24}$$

where λ_k is the Lebesgue measure on $I\!R^k$ and therefore:

$$\forall\, \phi \in \Phi_1^{iid}: \quad r(\Sigma_\phi) = k \tag{25}$$

In the second case, the distributions of Z_i's are considered as mixtures, the components of which have singular covariance matrix such that for any $a \in (0, 1)$, the mixed distributions have full rank covariance matrix.

$$\begin{aligned} \Phi_2^{iid} \ &= \ \{\phi \mid \phi = a\,\phi_1 + (1-a)\,\phi_2,\ a \in (0,1) \text{ with} \\ &\qquad r(\Sigma_\phi) = k \text{ and } r(\Sigma_{\phi_i}) < k,\ \forall\, i \in \{1,2\}\} \end{aligned} \tag{26}$$

In the third case, the matrix $\mathbf{Z}$ is treated as if it were not random, but with full rank, more explicitly, we consider:

$$\Phi_3^{nr} = \{\delta_{\mathbf{z}} \mid \ \mathbf{z} \in \mathbf{R}^*_{\mathbf{Z}} = \{\mathbf{z} \in I\!R^{n(k+1)} : r(\mathbf{z}) = k+1\}\} \tag{27}$$

If $\Phi = \Phi_1^{iid}$, then the regression model is *almost surely* identified and therefore identified for any finite sample size $n \geq k+1$, because (25) implies:

$$\forall\, n \geq k+1, \forall\, \phi \in \Phi_1^{iid}: \quad \phi(\{\mathbf{z}_2 \mid \ r(\mathbf{z}_2) = k\}) = 1 \tag{28}$$

If $\Phi = \Phi_2^{iid}$, then for any *finite* sample size $n \geq k+1$, the regression model will be identified, but not *almost surely* identified, because:

$$\forall\, n \geq k+1, \forall\, \phi \in \Phi_2^{iid} : \ \phi(\{\mathbf{z}_2 \mid \ r(\mathbf{z}_2) < k\}) = a^n + (1-a)^n > 0 \tag{29}$$

We nevertheless have the *almost sure* identification in the *asymptotic* model.

Finally, if $\Phi = \Phi_3^{nr}$, the regression model will be $\mathbf{R}^*_{\mathbf{Z}}$-*uniformly* identified. In this case, the definition of Φ_3^{nr} involves $n \geq k+1$.

Example 3

A particular case of example 2 may be illuminating, namely the case where the conditioning variable is univariate ($k = 1$) and binary. More specifically, let $Y_i \,|\, Z_i \sim I.N(\mu_{Z_i}\,, \sigma^2)$, be a conditional model such that $\forall\, i \in \{1, 2, ..., n\}$: Z_i are $i.i.d$ binary random variables, *i.e.*

$$P(Z_i = z \,|\, \phi) = \phi^z (1-\phi)^{1-z} \mathbf{I}_{\{0,1\}}(z) \;\text{ with }\; \phi \in \Phi \subset [0,1]$$

Following the above notation we have $(Y|\mathbf{Z}, \beta, \sigma^2) \sim N(\mathbf{Z}\beta, \sigma^2 I_n)$ where $\beta = (\mu_0\,, \mu_1 - \mu_0)'$ and $\mathbf{Z} = (i_n, Z)$ with $i_n = (1, 1, ..., 1)' \in I\!R^n$ and $Z = (Z_1, Z_2, ..., Z_n)' \in \{0,1\}^n$.
If $\phi = 0$ (respectively $\phi = 1$), μ_1 (respectively μ_0) would not be identified. Thus a sufficient condition for the identification of this model is $\Phi = (0, 1)$. Note however that in finite sample size, this model is not *almost surely* identified because even for $\phi \in (0,1)$, the probability of $Z = \mathbf{0}$ (respectively $Z = i_n$) is equal to $(1-\phi)^n > 0$ (respectively $\phi^n > 0$). Nevertheless, under $\Phi = (0, 1)$, the asymptotic model is *almost surely* identified.

4.2 Identification, unbiasedness and consistency

Let us now consider further the identification for finite sample size n (resp. in the asymptotic model) as necessary condition for the existence of an unbiased (resp. consistent) estimator.

Similarly to the weak identification, weak unbiasedness of T does not prevent that for any $z \in \mathbf{R}_Z$, T would not be unbiased in the submodel $\mathbf{M}^{\theta}_{Y|z} = \{(\mathbf{R}_Y, \mathcal{Y}), \{P_Y^{z,\theta} : \;\theta \in \Theta\}\}$. When such a feature is considered undesirable, one may strengthen the weak condition of unbiasedness in a way similar to what has been suggested in definition 6, for the identification in conditional model. For this purpose, let $g(\theta)$ be the parameter of interest and $T = f(Y, Z)$ an estimator of $g(\theta)$ and denote:

$$C_{g,T}(\theta) = \{z \in \mathbf{R}_Z \mid \; I\!E^{\theta}(T \mid Z = z) = g(\theta)\} \tag{30}$$

Let us also consider:

$$\begin{aligned} C_{g,T} &= \bigcap_{\theta\in\Theta} C_{g,T}(\theta) && (31)\\ &= \{z\in\mathbf{R}_Z \mid \ \forall\,\theta\in\Theta : I\!\!E^{\theta}(T \mid Z=z) = g(\theta)\} && (32)\end{aligned}$$

Definition 7 *Let us assume that $C_{g,T}$ is $\mathcal{Z}$-measurable. Then, T is said to be unbiased (resp. almost surely unbiased) for $g(\theta)$ if:*

$$\forall\,\phi\in\Phi:\ \ \phi(C_{g,T}) > 0 \ (resp. = 1) \tag{33}$$

T is said to be uniformly unbiased for $g(\theta)$ if:

$$\forall\,\theta\in\Theta,\,\forall\,z\in\mathbf{R}_Z:\ \ I\!\!E^{\theta}(T \mid Z=z) = g(\theta) \tag{34}$$

■

Clearly, *uniform* unbiasedness implies *almost sure* unbiasedness which implies *unbiasedness* wich also implies the *weak* one. The *almost sure* unbiasedness of T_n for $g(\theta)$ requires the set of $z\in\mathbf{R}_Z$, for wich T is unbiased for $g(\theta)$ in the model $\mathbf{M}^{\theta}_{Y|z} = \{(\mathbf{R}_Y,\mathcal{Y}),\{P^{z,\theta}_Y :\ \ \theta\in\Theta\}\}$, to have probability one for every $\phi^n\in\Phi$.

Theorem 5 *A necessary condition for the existence of an unbiased, almost surely unbiased or uniformly unbiased estimator of $g(\theta)$ is that $g(\theta)$ be respectively identified, almost surely identified or uniformly identified.*

The same argument as above is used to define three other levels of consistency. Let us first define:

$$D_{(T_n),g}(\theta) = \{z\in\mathbf{R}_Z \mid \ \forall\,\epsilon>0 : \lim_{n\longrightarrow+\infty} P^{z,\theta}_Y(|T_n - g(\theta)| < \epsilon) = 1\} \tag{35}$$

Let us also consider:

$$\begin{aligned} D_{(T_n),g} &= \bigcap_{\theta\in\Theta} D_{(T_n),g}(\theta) && (36)\\ &= \{z\in\mathbf{R}_Z \mid \ \forall\,\theta\in\Theta,\forall\,\epsilon>0 : \lim_{n\longrightarrow+\infty} P^{z,\theta}_Y(|T_n - g(\theta)| < \epsilon) = 1\}\end{aligned}$$

Definition 8 *Let us assume that $D_{(T_n),g}$ is $\mathcal{Z}$-measurable. Then, the sequence of estimators T_n is said to be consistent (resp. almost surely consistent) if:*

$$\forall\, \phi \in \Phi : \ \phi(D_{(T_n),g}) > 0 \ (resp. = 1) \tag{37}$$

The sequence of estimators T_n is said to be uniformly consistent if:

$$\forall\, z \in \mathbf{R}_Z, \forall\, \theta \in \Theta, \forall\, \epsilon > 0 : \lim_{n \longrightarrow +\infty} P_Y^{z,\theta}(|T_n - g(\theta)| < \epsilon) = 1 \tag{38}$$

■

Next theorem, similar to theorem 5, shows that the necessity of identification for the existence of a consistent estimator depend on the level of consistency that is required.

Theorem 6 *A necessary condition for the existence of a consistent, almost surely consistent or uniformly consistent estimator of $g(\theta)$ is that $g(\theta)$ be respectively identified, almost surely identified or uniformly identified in the asymptotic model.*

5 Conclusion

In this paper, we consider that a *conditional* model is a statistical model bearing on all the variables, *i.e.* on the "endogenous variables", Y, and the conditioning, or "exogenous", variables, Z, such that ϕ, the parameter characterizing the marginal distribution of Z, is a nuisance parameter that is identified and "well-separated" from θ, the parameter of interest, characterizing the Z-conditional distribution. Therefore, a conditional model is characterised by a family of marginal distributions on the exogenous variables *and* a family of "well specified" transitions of probabilities, playing a role of conditional probabilities in a global model. Typically, but not always, ϕ takes values in a "thick" subset, Φ, of all the probability distributions of Z.

By taking explicitely into account the randomness of Z, we show that the identification in conditional models may crucially depend on conditions on Φ. Furthermore, four levels of identification in conditional models, stress different intensities of this dependence on Φ: the weak, the standard, the almost sure and the uniform one. The weak one, frequently used in the statistical literature, is related to the joint distribution on $X = (Y, Z)$. The unique "conditional aspect" of this concept is that $g(\theta)$ is a function of the parameter θ characterizing the conditional distribution. The next two concepts *i.e.* the standard, respectively the almost sure one, makes explicitely use of the conditional distribution. They require the set of z values, for which the parameter of interest $g(\theta)$ is identified from the conditional distribution, to be a ϕ-non nul subset of $\mathbf{R}_Z$, respectively of measure one. Finally, the uniform concept, the strongest one, assumes that the parameter function $g(\theta)$ is identified from $P_Y^{z,\theta}$, for every $z \in \mathbf{R}_Z$. In this last concept, the exogenous variables are treated as if they were not random. Other concepts such as consistency and unbiasedness in conditional models, have been shown to follow the same structure as identification. As a final conclusion, the table below exhibits the relationships between all these concepts.

Relationships between the four levels of identification, unbiasedness and consistency.

Unbiasedness		**Identification**		**Consistency**
		usual in stat. literature		
		Theorem 1 $\Updownarrow$		
Weak eq. (12)	Theorem 2 $\Longrightarrow$	Weak eq. (8)	Theorem 3 $\Longleftarrow$	Weak eq. (14)
$\Uparrow$		$\Uparrow$		$\Uparrow$
Standard eq. (33)	Theorem 5 $\Longrightarrow$	Standard definition 6, part. a)	Theorem 6 $\Longleftarrow$	Standard eq. (37)
$\Uparrow$		$\Uparrow$		$\Uparrow$
Almost sure eq. (33)	Theorem 5 $\Longrightarrow$	Almost sure definition 6, part. b)	Theorem 6 $\Longleftarrow$	Almost sure eq. (37)
$\Uparrow$		$\Uparrow$		$\Uparrow$
Uniform eq. (34)	Theorem 5 $\Longrightarrow$	Uniform definition 6, part. c)	Theorem 6 $\Longleftarrow$	Uniform eq. (38)

eq: *Equation number in the text.*
$\Uparrow$: *Trivial.*
$\Longrightarrow$ and $\Longleftarrow$: *Theorem or Proposition.*

Appendix:

Proof of Lemma 1

We need only to prove that (10) implies (11) because the other implication is trivial. Note first that (10) is equivalent to:

$$\forall (A,B) \in \mathcal{Y} \times \mathcal{Z} : \int_B P_Y^{z,\theta_1}(A) d\phi = \int_B P_Y^{z,\theta_2}(A) d\phi$$

This implies:

$$\forall A \in \mathcal{Y} : \; P_Y^{Z,\theta_1}(A) = P_Y^{Z,\theta_2}(A) \;\; \phi - a.s$$

Or equivalently:

$$\forall A \in \mathcal{Y}, \, \exists \, N_A \in \mathcal{Z} : \, \phi(N_A) = 0 \, \text{and} \, \forall z \in N_A^c : \; P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A) \quad (39)$$

Let $\mathcal{Y}_0$ be a countable π-system generating $\mathcal{Y}$ $i.e.$ $\mathcal{Y} = \sigma(\mathcal{Y}_0)$ and define a subset $N = \bigcup_{A \in \mathcal{Y}_0} N_A$, with N_A as defined in (39). Clearly, $\phi(N) = 0$, furthermore:

$$\forall A \in \mathcal{Y}_0, \; \forall z \in N^c : \; P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)$$

wich implies by a monotone class theorem:

$$\forall A \in \mathcal{Y}, \; \forall z \in N^c : \; P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)$$

Therefore, condition (39) and the separability of $\mathcal{Y}$ imply:

$$\exists \, N \in \mathcal{Z}, \; \phi(N) = 0, \; \forall \, z \in N^c : \; P_Y^{z,\theta_1} = P_Y^{z,\theta_2}$$

which is equivalent to (11).

$\blacksquare$

Proof of Theorem 1

The Proof of (i) :

For every $(\theta_1, \theta_2) \in \Theta^2$, let $B(\theta_1, \theta_2) = \{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}$, then the separability of $\mathcal{Y}$ ensures the measurability of $B(\theta_1, \theta_2)$, more specifically:

$$\begin{aligned}\left\{z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\right\} &= \left\{z \mid P_Y^{z,\theta_1} = P_Y^{z,\theta_2}\right\}^c \\ &= \left(\bigcap_{A\in\mathcal{Y}_0} \left\{z \mid P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)\right\}\right)^c \\ &= \bigcup_{A\in\mathcal{Y}_0} \left\{z \mid P_Y^{z,\theta_1}(A) \neq P_Y^{z,\theta_2}(A)\right\} \in \mathcal{Z} \qquad (40)\end{aligned}$$

The proof of (ii):

By lemma 1 and property (i) of theorem 1, we have:

$$\begin{aligned}\phi \otimes P_Y^{Z,\theta_1} \neq \phi \otimes P_Y^{Z,\theta_2} &\iff \phi(B(\theta_1, \theta_2)^c) < 1 \\ &\iff \phi(B(\theta_1, \theta_2)) > 0 \qquad (41)\end{aligned}$$

■

Proof of Theorem 3

Let us prove theorem 3 by contradiction. Indeed, suppose that $g(\theta)$ is not weakly identified in the asymptotic model; This means that there exists $(\phi^0, \theta_1^0, \theta_2^0) \in \Phi \times \Theta_g$ such that $\phi^0 \otimes P_Y^{Z,\theta_1^0} = \phi^0 \otimes P_Y^{Z,\theta_2^0}$.

Suppose that there also exists a consistent estimator of $g(\theta)$, say T_n. This means, by definition 4, that:

$$\forall\, \omega = (\theta, \phi) \in \Theta \times \Phi : \; p^{\omega} lim\, T_n = g(\theta) \qquad (42)$$

Let now $\omega_i^0 = (\theta_i^0, \phi^0)$ with $i \in \{1, 2\}$. By equation (42), $p^{\omega_i^0} lim\, T_n = g(\theta_i^0)$ for $i \in \{1, 2\}$. This implies that $g(\theta_1^0) = g(\theta_2^0)$ because, by hypothesis, $P^{\omega_1^0} = P^{\omega_2^0}$, and this contradicts the fact that $(\theta_1^0, \theta_2^0) \in \Theta_g$.

■

Proof of Theorem 4

To prove theorem 4, it is sufficient to prove that (15) implies (16), the converse implication being trivial. This is equivalent to prove that the negation of (16) implies the negation of (15) *i.e.* if there exist parameter values $(\theta_1, \phi^{(n)*})$ and $(\theta_2, \phi^{(n)*})$ for wich the model is not identified for a sample of size n, then there exist parameter values $(\theta_1, \phi^{(n+1)*})$ and $(\theta_2, \phi^{(n+1)*})$ for wich this model is not identified for a sample of size $n+1$.

For a fixed n, let us assume under condition (C.2) that there exist parameter values $(\theta_1, \phi^{(n)*})$ and $(\theta_2, \phi^{(n)*})$ such that $g(\theta_1) \neq g(\theta_2)$ and:

$$\phi^{(n)*} \otimes P_{Y_1^n}^{Z_1^n,\theta_1} = \phi^{(n)*} \otimes P_{Y_1^n}^{Z_1^n,\theta_2} \tag{43}$$

This is equivalent to:

$$\forall (A,B) \in \mathcal{Y}_1^n \otimes \mathcal{Z}_1^n : \int_B P_{Y_1^n}^{z_1^n,\theta_1}(A) d\phi^{(n)*} = \int_B P_{Y_1^n}^{z_1^n,\theta_2}(A) d\phi^{(n)*} \tag{44}$$

If the model is not identified for sample of size n, then it is not identified for samples of sizes less than n. Furthermore, in the sifted case, the parameter values (θ_1, θ_2) for wich the model is not identified for sample of size n are the same as that ones for wich this model is not identified for samples of sizes less than n, in particular for $n = 1$. Therefore, using assumption (C.1), equation (44) implies that there exists $\phi^{(1)*} \in \Phi^{(1)}$ such that:

$$\forall (A,B) \in \mathcal{Y}_1 \otimes \mathcal{Z}_1 : \int_B P_{Y_1}^{z_1,\theta_1}(A) d\phi^{(1)*} = \int_B P_{Y_1}^{z_1,\theta_2}(A) d\phi^{(1)*} \tag{45}$$

By assumption (C.4), there exists a transition $\phi_{n+1}^{n*} \in \Phi_{n+1}^n$ such that:

$$\forall\, \phi^{(n)} \in \Phi^{(n)} : \ \phi^{(n)}\Big\{ z_1^n \mid \forall B \in \mathcal{Z}_1 : \ \phi_{n+1}^{n*}(B \mid z_1^n) = \phi^{(1)*}(B) \Big\} = 1$$

Under condition (C.3), we may consider $\phi^{(n+1)*} = \phi^{(n)*} \otimes \phi_{n+1}^{n*}$ and show that, under the sifting condition, the probability measures corresponding to $\omega^{(n+1)*} = (\theta_1, \phi^{(n+1)*})$ and $\omega^{(n+1)**} = (\theta_2, \phi^{(n+1)*})$ coincide; indeed:

$\forall\, (A,B) \in \mathcal{Y}_1^{n+1} \otimes \mathcal{Z}_1^{n+1}$ such that $A = A^n \times A_{n+1} \in \mathcal{Y}_1^n \times \mathcal{Y}_1$ and $B = B^n \times B_{n+1} \in \mathcal{Z}_1^n \times \mathcal{Z}_1$:

$$\int_B P_{Y_1^{n+1}}^{z_1^{n+1},\theta_1}(A) d\phi^{(n+1)*} \quad \overset{\text{C.5}}{=} \quad \int_B P_{Y_1^n}^{z_1^n,\theta_1}(A^n) P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi^{(n+1)*}$$

$$
\begin{aligned}
&= \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n)[\int_{B_{n+1}} P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi_{n+1}^{n*}]\ d\phi^{(n)*} \\
&\overset{C.4}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n) d\phi^{(n)*} \int_{B_{n+1}} P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi^{(1)*} \\
&\overset{C.6}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n) d\phi^{(n)*} \int_{B_{n+1}} P_{Y_1}^{z_1,\theta_1}(A_{n+1})\ d\phi^{(1)*} \\
&\overset{(44)\,\&\,(45)}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_2}(A^n) d\phi^{(n)*} \int_{B_{n+1}} P_{Y_1}^{z_1,\theta_2}(A_{n+1})\ d\phi^{(1)*} \\
&= \int_{B} P_{Y_1^{n+1}}^{z_1^{n+1},\theta_2}(A)\ d\phi^{(n+1)*}
\end{aligned}
$$

Proof of Theorem 5

Let T be an *unbiased* (resp. *almost surely unbiased*) estimator of $g(\theta)$. This implies that $\forall\, \phi \in \Phi$:

$$\phi(\{z \mid \forall\, \theta \in \Theta :\ \mathbb{E}^{\theta}(T \mid z) = g(\theta)\}) > 0\ (= 1) \tag{46}$$

Wich also implies that $\forall\, \phi \in \Phi$:

$$\phi(\{z \mid \forall\, (\theta_1, \theta_2) \in \Theta_g :\ \mathbb{E}^{\theta_1}(T \mid z) \neq \mathbb{E}^{\theta_2}(T \mid z)\}) > 0\ (= 1)$$

Wich also implies that $\forall\, \phi \in \Phi$:

$$\phi(\{z \mid\ \forall\, (\theta_1, \theta_2) \in \Theta_g :\ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}) > 0\ (= 1) \tag{47}$$

Equation (47) means that the parameter $g(\theta)$ is identified (resp. almost surely identified).

■

Proof of Theorem 6

We will prove that identification in the asymptotic model is a necessary condition for the existence of a consistent estimator. For the almost sure and the uniform cases, the proof is similar. We also make use of the sets $R_{Z,g}$ and $D_{(T_n),g}$ respectively defined in (19) and (36), and assumed to be measurable in the definitions 6 and 8.

Let us assume that $g(\theta)$ is not identified in the asymptotic model $\mathbf{M}_Y^{Z,\theta}$. This implies, by definition 6, that there exists a law, say $\phi^0 \in \Phi$, on the space

$(\mathbf{R}_X, \mathcal{X})$ of the stochastic process X, such that $\phi^0(R_{Z,g}) = 0$ or equivalently that $\phi^0(R^c_{Z,g}) = 1$, where:

$$R^c_{Z,g} = \{z \in \mathbf{R}_Z \mid \exists\, (\theta^z_1, \theta^z_2) \in \Theta_g :\ P_Y^{z,\theta^z_1} = P_Y^{z,\theta^z_2}\} \tag{48}$$

Now, let T_n be a consistent estimator of $g(\theta)$. This implies, by definition 8, that $\phi^0(D_{(T_n),g}) > 0$. Consequently, $\phi^0(D_{(T_n),g} \cap R^c_{Z,g}) > 0$. Let us show, however, that $D_{(T_n),g} \cap R^c_{Z,g} = \emptyset$. Indeed, let $z \in D_{(T_n),g} \cap R^c_{Z,g}$. That $z \in R^c_{Z,g}$ means that there exists $(\theta^z_1, \theta^z_2) \in \Theta_g$ such that $P_Y^{z,\theta^z_1} = P_Y^{z,\theta^z_2} = P_Y^{z,0}$, say. Furthermore, that $z \in D_{(T_n),g}$ means that:

$$\forall\, \epsilon > 0, \forall\, i \in \{1, 2\} :\ \lim_{n \longrightarrow +\infty} P_Y^{z,0}(|T_n - g(\theta^z_i)| < \epsilon) = 1 \tag{49}$$

So, T_n converges according to the same probability measure $P_Y^{z,0}$ to both $g(\theta^z_1)$ and $g(\theta^z_2)$. This implies that $g(\theta^z_1) = g(\theta^z_2)$, in contradiction with the non identification of $g(\theta)$ in the asymptotic model.

■

Aknowledgements. *Without implications, the authors gratefully acknowledge useful comments on earlier versions by L. Bauwens, C. Almeida, Steven Wei and P. Wilson. The authors are particularly indebted to J- M. Rolin for many discussions and comments that led to substantially improve some results as well as the general exposition.*

References

BUNKE, H. and O.BUNKE, (1974), Identification and Estimability, *Mathematische Operationsforschung und Statistik*, **5**, 223-233.

FLORENS, J.-P., M.MOUCHART and J.-M. ROLIN, (1985), On two Definitions of Identification, *Statistics*, **2**, 213-219.

FLORENS, J.-P., M.MOUCHART and J.-M.ROLIN, (1990), *Elements of Bayesian Statistics*, Marcel Dekker.

KOOPMANS, T. C. and O.REIERSØL, (1950), The Identification of Structural Characteristics, *The Annals Of Mathematical Statistics*, 165-181.

MANSKI, C. F., (1988), Identification in Binary Response models, *Journal of the American Statistical Association*, Vol. 83, 729-738.

MATZKIN, R. L., (1993), Nonparametric identification and estimation of polychotomous choice models, *Journal of Econometrics*, Vol. 58, 137-168.

MOUCHART, M., J.-M.ROLIN and E.SCHEILING, (1997), Bayesian Identification of Semi-Parametric Binary Response Models, Discussion Paper 9713, Institut de statistique, Louvain-la-Neuve (B).

ROEHRIG, C. S., (1988), Conditions for Identification in Nonparametric and Parametric Models, *Econometrica*, Vol. 56, No. 2, 433-447.